[S11-4]

Multi-Omics-Based Systems Approach Captures the Whole-Cellular Features of Closely Related *Escherichia coli* B and K-12

CSung Ho Yoon[1]*, Mee-Jung Han[2], Haeyoung Jeong[1], Choong Hoon Lee[3], Xiao-Xia Xia[2], Ji Hoon Shim[1], Sang Yup Lee[2,4], Tae Kwang Oh[5], and Jihyun F. Kim[1,6]

[1]Systems Microbiology Research Center, and [5]21C Frontier Microbial Genomics & Applications Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, [2]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering (BK21 Program), BioProcess Engineering Research Center, Center for Systems and Synthetic Biotechnology, and Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, [3]Department of Biological Sciences, KAIST, [4]Department of Bio and Brain Engineering, and Bioinformatics Research Center, KAIST, [6]School of Science, Korea University of Science and Technology (UST), Daejeon 305-333

Recent advances in omics technologies provide us with the possibility of deciphering an organism's genotype-to-phenotype relationships. Starting from the genome sequence, analyses of transcriptome and proteome data yield information on global expression of genes and proteins. Phenotype microarray (PM) technology offers a simultaneous monitoring of nearly 2,000 phenotypic data (1). An integrative analysis of the system-level measurements can complement to each other to reveal deeper insights into biological systems (2, 3). The multidimensional omics data can be further integrated to reconstruct genome-wide computational models, which can generate testable hypotheses on cellular function (4). However, as the analysis of individual genome may fail to identify important features, comparing various aspects of closely related strains can be important in understanding the cellular inner workings and variations among organisms.

Escherichia coli is one of the best studied organisms and has been widely employed in scientific studies as well as industrial applications. Almost all laboratory strains of *E. coli* are derivatives of non-pathogenic K-12 or B strains. Most genetic and metabolic studies have been performed with K-12 or its derivatives, which have been accelerated by the availability of the complete genome sequences of the MG1655 and W3110 strains. Recently, genome sequencing of two *E. coli* strains of the B lineage, REL606 and BL21(DE3), have been performed by International *E. coli* B Genome Consortium (5). Although the genome sequences of B and K-12 are highly similar based on comparison of IS elements (6), B strains often show phenotypes distinct from those of K-12 (7). Derivatives of *E. coli* B, especially BL21, have

been widely used for the overproduction of recombinant proteins, ethanol, and other biomolecules (8), by virtue of several favorable features including faster growth in minimal media, lower acetate production, higher expression levels of recombinant proteins, and less degradation of such proteins during purification (9). Here, we systematically integrated the comparative results of the genomes, transcriptomes, proteomes, and phenomes of *E. coli* B REL606 and K-12 MG1655 to decipher the organism-wide characteristics that differentiate the two sequenced organisms. Also, proteome profiles of BL21(DE3) and W3110 were additionally analyzed for more comprehensive comparison of B and K-12.

Through integration of multidimensional omics data, we identified characteristic features of the B strains as follows. Lack of flagellar biosynthetic genes and low expression of motility-related genes can be important properties when B is used as a cell factory, because flagellar biosynthesis is energy-intensive and is not necessary under an industrial setup of constant agitation and generous supply of nutrients. Differences in the composition of the LPS core and expression of outer membrane proteins may influence the permeability and integrity of the cell envelope, which presumably results in alterations to screening barriers that control import and export of materials such as antibiotics, nutrients, and proteins. Importantly, the existence of the second T2S secretion system and enhanced capability for protein release suggest that B might be better suited for extracellular production of recombinant proteins. B exhibited up-regulation of amino acid biosynthetic genes, and showed lower expression of proteases. These characteristics, either coincidental or carefully considered when BL21 was chosen for strain engineering, are presumably desirable for the enhanced production of recombinant proteins.

A genome-scale metabolic network model of REL606 was reconstructed by incorporating genetic differences from MG1655 (10). Flux balance analysis was used to predict cell growth on various substrates, and the results were compared with the results from PM tests (4). *In silico* complementation experiments of the reconstructed model verified 23 reactions responsible for the phenotypes unique to REL606.

In summary, differences in utilization of a variety of nutrient substrates were identified and in many cases could be correlated with genotypic and functional differences. Also, transcriptome and proteome data cross-validated information on differentially expressed genes and proteins, demonstrating the utility of integrating omics data collected at various layers from genome to phenome. Genome-scale in silico models can generate and validate hypotheses on cellular processes. The integrated omics information described here on the cellular metabolism and physiology of *E. coli* B and K-12 should be pivotal in better understanding the underlying biological networks and is invaluable for designing strains having customized genomes as well as establishing rational fermentation strategies. It is also expected that the omics-based systems approaches reported here will provide a framework for elucidating the phenetic characteristics of other organisms whose genomes are sequenced.

Acknowledgements

This work was supported by the 21C Frontier Microbial Genomics and Applications Center Program (to J.F.K.), and a Korean Systems Biology Research Grant (to S.Y.L.) of the Ministry of Science and Technology. A special gratitude also goes to the 21C Frontier Microbial Genomics and Applications Center for financial support for fabrication of DNA microarrays.

References

1. BR Bochner, *Nat. Rev. Genet.*, **4**, 309, 2003.
2. SH Yoon, MJ Han, SY Lee, KJ Jeong, and JS Yoo, *Biotechnol. Bioeng.*, **81**, 753, 2003.
3. T Ideker et al., *Science*, **292**, 929, 2001.
4. MW Covert, EM Knight, JL Reed, MJ Herrgard, and BO Palsson, *Nature*, **429**, 92, 2004.
5. H Jeong et al., submitted, 2007.
6. D Schneider et al., *BMC Microbiol.*, **2**, 18, 2002.
7. JR Swartz, in *Escherichia coli* and *Salmonella*: cellular and molecular biology, F. C. Neidhardt et al., Eds. (ASM Press, Washington, D.C., 1996), pp. 1693-1712.
8. JH Choi, KC Keum, and SY Lee, *Chem. Eng. Sci.*, **61**, 876, 2006.
9. GW Luli and WR Strohl, *Appl. Environ. Microbiol.*, **56**, 1004, 1990.
10. AM Feist et al., *Mol. Syst. Biol.*, **3**, 121, 2007.